Average-Case Computational Complexity of Tensor Decomposition

Alex Wein

University of California, Davis

Department of Mathematics

UCDAVIS

Tensor Decomposition

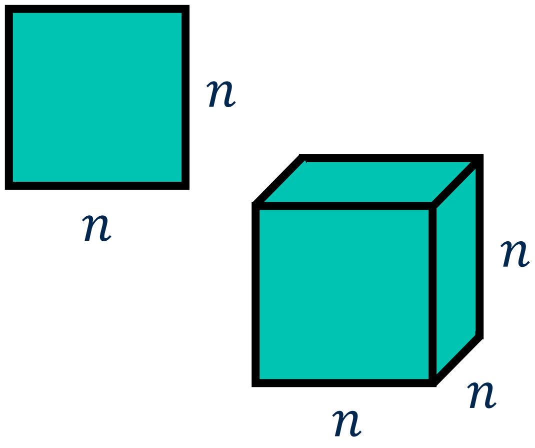
Basic algorithmic primitive with applications in:

- Phylogenetic reconstruction [MR05]
- Topic modeling [AFHKL12]
- Community detection [AGHK13,HS17,AAA17,JLLX20]
- Learning Gaussian mixtures [HK13,GHK15,BCMV14,ABGRV14]
- Independent component analysis [GVX14]
- Dictionary learning [BKS15,MSS16]
- Multi-reference alignment [PWBRS19]
- …

Tensors

Order-2 tensor: matrix $M = (M_{ij})$

Order-3 tensor $T = (T_{ijk})$



Rank-1 (symmetric) order-2 tensor $vv^\top$ $(vv^\top)_{ij} = v_i v_j$ $v \in \mathbb{R}^n$

Rank-1 (symmetric) order-3 tensor $v^{\otimes 3}$ $\left(v^{\otimes 3}\right)_{ijk} = v_i v_j v_k$

Random Tensor Decomposition

Given a rank-r order-3 tensor

$$T = \sum_{i=1}^{r} a_i^{\otimes 3} \qquad a_i \in \mathbb{R}^n$$

the goal is to recover the components $a_1, \ldots, a_r$

Assume random components $a_i \sim \mathcal{N}(0, I_n)$

succeed with high probability

method of moments

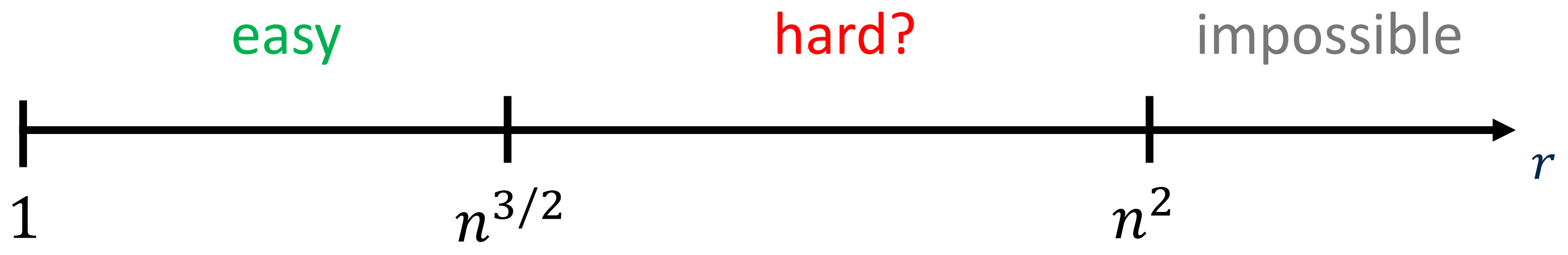
$$T = \sum_{i=1}^{r} a_i^{\otimes 3} \qquad a_i \sim \mathcal{N}(0, I_n)$$

Prior Work

Algorithmic results: SoS [GM15, Ma-Shi-Steurer'16], spectral [HSSS16,DOLST22], …

All known poly-time algorithms require $r \ll n^{3/2}$ $\ll$ hides polylog factor

Information-theoretically possible when $r \leq cn^2$ [BCO14] c = constant



Q: is this hardness inherent?

Statistical-Computational Gaps

Many statistical problems have “hard” regimes

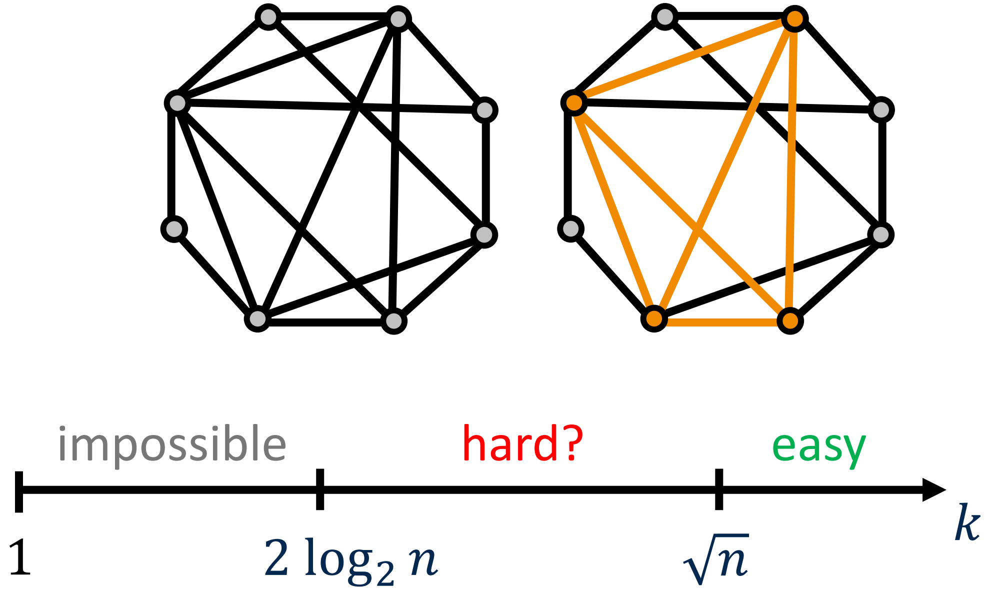
sparse PCA, compressed sensing, community detection, tensor PCA, …

No average-case complexity theory

Instead:

- Reductions from planted clique
- Lower bounds in restricted models
- Optimization landscape

Planted clique: G(n,1/2) + {k-clique}



Tensor Decomposition: Difficulties

Which lower bound framework?

- Reduction – out of reach?
- Statistical query (SQ) model – not applicable (no iid samples)
- Sum-of-squares (SoS) – hardness of refutation [BBKMW21]
- Optimization landscape – what function to optimize? [GZ19, BGJ20, CMZ22]
- Low-degree polynomials (LDP) – this talk

Tensor Decomposition: More Difficulties

Issue of symmetry

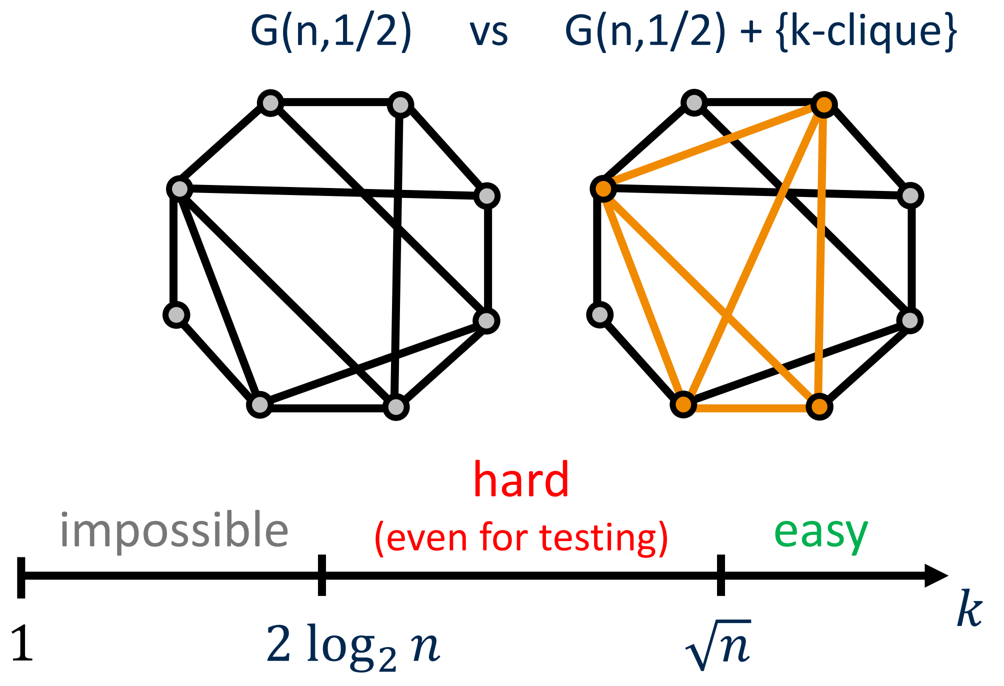
which component to recover?

$$T = \sum_{i=1}^{r} a_i^{\otimes 3}$$

Existing SQ/SoS/LDP lower bounds leverage hardness of testing vs iid "null"

a few exceptions [Schramm-**W**'22, Koehler-Mossel'21]

- Testing rank-r tensor vs iid tensor is easy when $r \ll n^3$
- But decomp hard when $r \gg n^{3/2}$



Solving the Issue of Symmetry

Define a new model: "largest component recovery"

$$T = (1+\delta)a_1^{\otimes 3} + \sum_{i=2}^{r} a_i^{\otimes 3} \qquad a_i \in \{\pm 1\}^n \text{ unif. at random}$$

relation to tensor PCA

Goal: recover/estimate $a_{11} \coloneqq (a_1)_1$

Hardness of the above problem implies hardness of decomposing

$$\sum_{i=1}^{r} \lambda_i a_i^{\otimes 3} \qquad \lambda_i \in [1, 1+\delta] \text{ arbitrary,} \qquad a_i \in \{\pm 1\}^n \text{ unif. at random}$$

Main Result: LDP Phase Transition

Class of algorithms: multivariate polynomials f in the entries of

$$T = (1+\delta)a_1^{\otimes 3} + \sum_{i=2}^{r} a_i^{\otimes 3} \qquad a_i \in \{\pm 1\}^n \text{ unif. at random}$$

Degree-D minimum mean squared error:

$$\mathrm{MMSE}_{\leq D} := \inf_{f \deg D} \mathbb{E}_a[(f(T) - a_{11})^2]$$

Theorem (W. '22) Fix any $\epsilon > 0,\ \delta > 0$

- *(Easy)* If $r \leq n^{3/2-\epsilon}$ then $\mathrm{MMSE}_{\leq O(\log n)} \to 0$ as $n \to \infty$
- *(Hard)* If $r \geq n^{3/2+\epsilon}$ then $\mathrm{MMSE}_{\leq n^{\Omega(1)}} \to 1$ as $n \to \infty$

Why LDP (Low-Deg Poly) Framework?

Algorithms captured by $O(\log n)$-deg poly: spectral, AMP, local, SQ, …

LDP lower bounds rule out certain known approaches

[BBHSL21]

Great track record of predicting stat-comp gaps

LDP lower bounds give rigorous "evidence" for hardness

- Some counterexamples: Gaussian elimination, lattice basis reduction, …
- But these algorithms tend to be "brittle"

Testing [Hopkins-Steurer'17, HKPRSS17, …], estimation [SW22], optimization [GJW20, …]

Connections to circuit complexity [Gamarnik-Jagannath-W'22]

Main Result: LDP Phase Transition

Class of algorithms: multivariate polynomials f in the entries of

$$T = (1+\delta)a_1^{\otimes 3} + \sum_{i=2}^{r} a_i^{\otimes 3} \qquad a_i \in \{\pm 1\}^n \text{ unif. at random}$$

Degree-D minimum mean squared error:

$$\mathrm{MMSE}_{\leq D} := \inf_{f \deg D} \mathbb{E}_a[(f(T) - a_{11})^2]$$

Theorem (W. '22) Fix any $\delta > 0$, $\epsilon > 0$

- *(Easy)* If $r \leq n^{3/2-\epsilon}$ then $\mathrm{MMSE}_{\leq O(\log n)} \to 0$ as $n \to \infty$
- *(Hard)* If $r \geq n^{3/2+\epsilon}$ then $\mathrm{MMSE}_{\leq n^{\Omega(1)}} \to 1$ as $n \to \infty$

Upper Bound: LDP Succeeds

Idea: “spectral methods from tensor networks”

[Hopkins-Schramm-Shi-Steurer’16, Moitra-**W**’19, Ding-d’Orsi-Liu-Steurer-Tiegel’22]

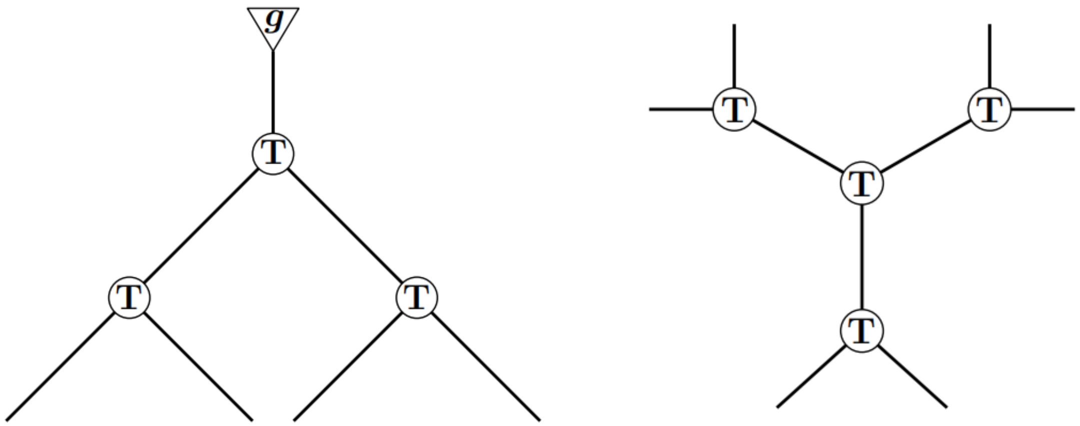
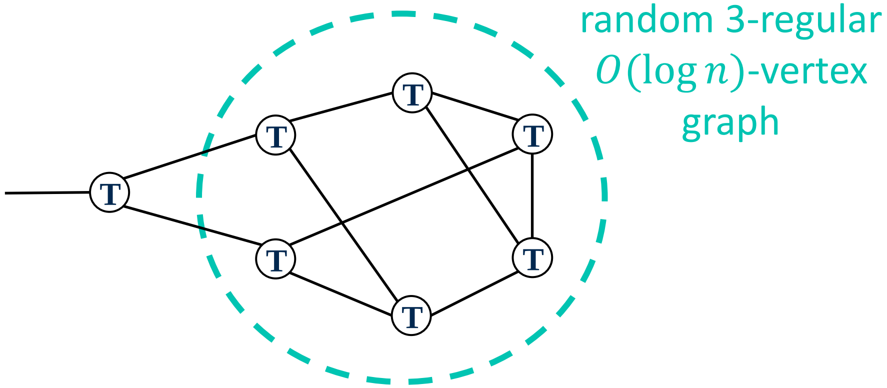


Image credit: [DOLST22]



This work

Degree-$O(\log n)$ polynomial implies quasipoly-time $n^{O(\log n)}$ algorithm

Main Result: LDP Phase Transition

Class of algorithms: multivariate polynomials f in the entries of

$$T = (1+\delta)a_1^{\otimes 3} + \sum_{i=2}^{r} a_i^{\otimes 3} \qquad a_i \in \{\pm 1\}^n \text{ unif. at random}$$

Degree-D minimum mean squared error:

$$\mathrm{MMSE}_{\leq D} := \inf_{f \deg D} \mathbb{E}_a[(f(T) - a_{11})^2]$$

Theorem (W. '22) Fix any $\delta > 0$, $\epsilon > 0$

- *(Easy)* If $r \leq n^{3/2-\epsilon}$ then $\mathrm{MMSE}_{\leq O(\log n)} \to 0$ as $n \to \infty$
- *(Hard)* If $r \geq n^{3/2+\epsilon}$ then $\mathrm{MMSE}_{\leq n^{\Omega(1)}} \to 1$ as $n \to \infty$

Lower Bound: Baby Example

Observe scalar $t = \sum_{i=1}^{r} a_i$ $a_i \in \{\pm 1\}$ unif. at random

Goal: estimate a_1

$$\sup_{\hat{f}} \frac{\langle c, \hat{f} \rangle}{\sqrt{\hat{f}^\top P \hat{f}}} = \sqrt{c^\top P^{-1} c}$$

$$\text{Corr}_{\le D} := \sup_{f \deg D} \frac{\mathbb{E}[f(t) a_1]}{\sqrt{\mathbb{E}[f(t)^2]}} = o(1)$$

Want to show $\text{Corr}_{\le D} := \sup_{f \deg D} \frac{\mathbb{E}[f(t) a_1]}{\sqrt{\mathbb{E}[f(t)^2]}} = o(1)$ $f(t) = \sum_{d=0}^{D} \hat{f}_d t^d$

First attempt:

- Numerator linear in $\hat{f}$ $\mathbb{E}[f(t) a_1] = \sum_{d=0}^{D} \hat{f}_d \underbrace{\mathbb{E}[t^d a_1]}_{c_d} =: \langle c, \hat{f} \rangle$
- Denominator quadratic in $\hat{f}$ $\mathbb{E}[f(t)^2] = \sum_{d,d'=0}^{D} \hat{f}_d \hat{f}_{d'} \underbrace{\mathbb{E}[t^d t^{d'}]}_{P_{d,d'}} =: \hat{f}^\top P \hat{f}$

$t = \sum_{i=1}^{r} a_i \quad a_i \in \{\pm 1\}$

Lower Bound: Baby Example

want $\frac{\mathbb{E}[f(t)a_1]}{\sqrt{\mathbb{E}[f(t)^2]}} \leq \cdots$

$$\sum_{d=0}^{D} \hat{f}_d t^d = f(t) = g(a) = \sum_{U \subseteq [r]} \hat{g}_U a^U \longleftarrow a^U \coloneqq \prod_{i \in U} a_i$$

orthonormal: $\mathbb{E}\left[a^U a^{U'}\right] = \mathbb{1}_{U=U'}$

Claim: $\mathbb{E}[f(t)^2] = \mathbb{E}[g(a)^2] = \|\hat{g}\|^2$

Claim: $\hat{g} = M\hat{f}$ for some matrix M

Claim: suffices to construct an explicit left-inverse M^+ s.t. $M^+ M = I$

$$\sup_f \frac{\mathbb{E}[f(t)a_1]}{\sqrt{\mathbb{E}[f(t)^2]}} = \sup_{\hat{f}} \frac{\langle c, \hat{f} \rangle}{\|M\hat{f}\|} = \sup_{\hat{f}} \frac{c^\top M^+ M \hat{f}}{\|M\hat{f}\|} \leq \sup_{\hat{g}} \frac{c^\top M^+ \hat{g}}{\|\hat{g}\|} = \|c^\top M^+\|$$

$\uparrow$ $\|\hat{g}\|$

$$t = \sum_{i=1}^{r} a_i \qquad a_i \in \{\pm 1\}$$

Constructing the Left-Inverse

$$\sum_{d=0}^{D} \hat{f}_d t^d = f(t) = g(a) = \sum_{U \subseteq [r]} \hat{g}_U a^U$$

Recall: $\hat{g} = M\hat{f}$ want M^+ s.t. $M^+ M = I$

In other words: $M^+ \hat{g} = \hat{f}$ whenever $\hat{g} = M\hat{f}$

In other words: given (valid) $\hat{g}$, recover $\hat{f}$

$t = a_1 + a_2 + a_3 + a_4$

↓

Proof by example: $g(a) = a_1 a_2 a_3 + a_1 a_2 a_4 - 2a_1 a_3 - 2a_3 a_4 + \cdots$ $f(t) = ??$

$r = 4, D = 3$

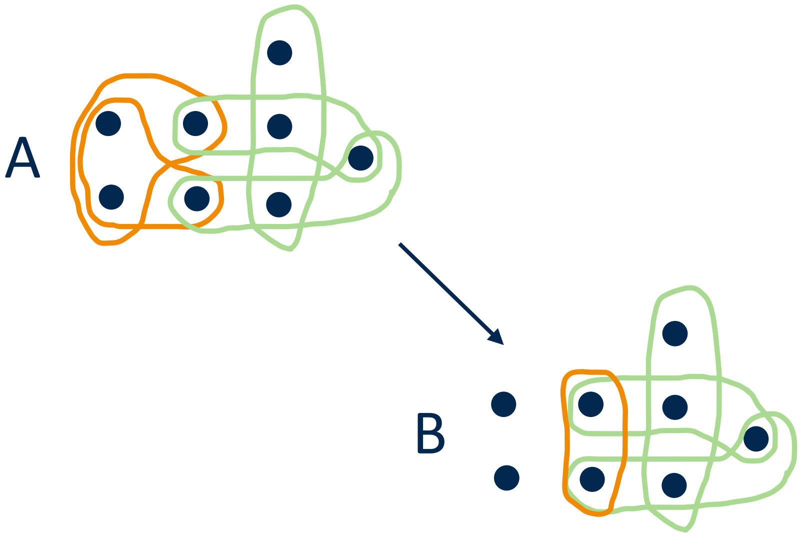
$$\frac{1}{6} t^3 = a_1 a_2 a_3 + a_1 a_2 a_4 + \cdots + \frac{5}{3}(a_1 + a_2 + a_3 + a_4)$$

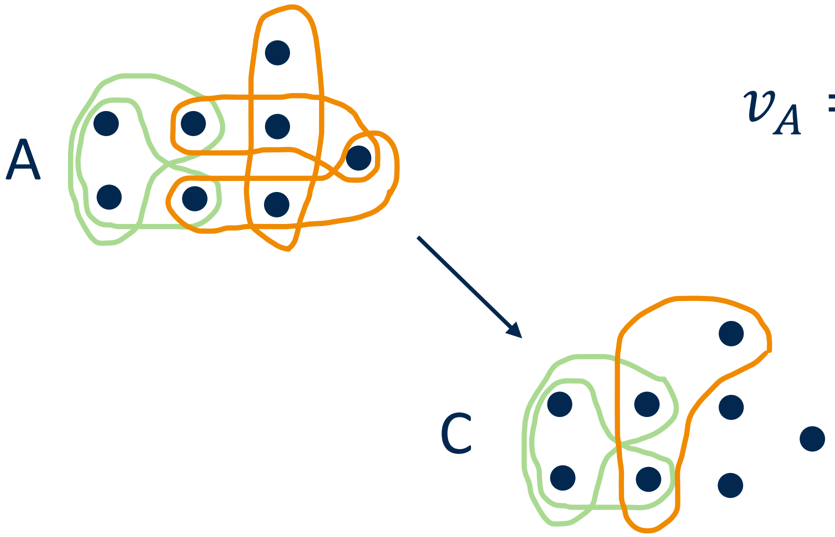
Wrapping Up the Lower Bound

Conclusion: $\mathrm{Corr}_{\leq D} \coloneqq \sup_{f \deg D} \frac{\mathbb{E}[f(t)a_1]}{\sqrt{\mathbb{E}[f(t)^2]}} \leq \|c^\top M^+\| =: \|v\|$

For the true model, v is indexed by hypergraphs and defined recursively

reminiscent of cumulants in [Schramm-**W**'22]





$$v_A = \alpha_{AB} v_B + \alpha_{AC} v_C + \cdots$$

Thanks!

Comments

First concrete lower bound for random tensor decomposition

low-degree polynomial threshold matches best known algorithms

Results extend to tensors of any order $k \geq 3$, threshold is $r \sim n^{k/2}$

Future directions: Gaussian components, structured tensors

Open: is "generic" tensor decomposition strictly harder than random ($k = 3$)?